A remote sensing derived dataset of 100 million individual tree crowns for the National Ecological Observatory Network

Ben. G. Weinstein[1] Sergio Marconi[1], Stephanie A. Bohlman[2], Alina Zare[3], Aditya Singh[4], Sarah J. Graves[5], Ethan P. White[1,6,7]

[1]Department of Wildlife Ecology and Conservation, University of Florida, Gainesville, Florida, USA

[2]School of Forest Resources and Conservation, University of Florida, Gainesville, Florida, USA

[3]Department of Electrical and Computer Engineering, University of Florida, Gainesville, Florida, USA

[4]Department of Agricultural & Biological Engineering, University of Florida, Gainesville, FL 32611, USA

[5] Nelson Institute for Environmental Studies, University of Wisconsin-Madison, Madison, Wisconsin, USA

[6]Informatics Institute, University of Florida, Gainesville, Florida, USA

[7]Biodiversity Institute, University of Florida, Gainesville, Florida, USA

Abstract

Forests provide biodiversity, ecosystem and economic services. Information on individual trees is important for understanding forest ecosystems but obtaining individual-level data at broad scales is challenging due to the costs and logistics of data collection. While advances in remote

sensing techniques allow surveys of individual trees at unprecedented extents, there remain technical challenges in turning sensor data into tangible information. Using deep learning methods, we produced an open-source dataset of individual-level crown estimates for 100 million trees at 37 sites across the United States surveyed by the National Ecological Observatory Network's Airborne Observation Platform. Each canopy tree crown is represented by a rectangular bounding box and includes information on the height, crown area, and spatial location of the tree. These data have the potential to drive significant expansion of individual-level research on trees by facilitating both regional analyses and cross-region comparisons encompassing forest types from most of the United States.

Introduction

Trees are central organisms in maintaining the ecological function, biodiversity and the health of the planet. There are estimated to be over three trillion individual trees on earth (Crowther et al., 2015) covering a broad range of environments and geography (Hansen et al., 2013). Counting and measuring trees is central to understanding key environmental and economic issues and has implications for global climate, land management and wood production. Field-based surveys of trees are generally conducted at local scales (~0.1-100 ha) with measurements of attributes for individual trees within plots collected manually. Connecting these local scale measurements at the plot level to broad scale patterns is challenging because of spatial heterogeneity in forests. Many of the central processes in forests, including change in forest structure and function in response to disturbances such as hurricanes and pest outbreaks, and human modification through forest management and fire, occur at scales beyond those feasible for direct field measurement.

Satellite data with continuous global coverage have been used to quantify important patterns in forest ecology and management such as global tree cover dynamics and disturbances in temperate forests (e.g. Bastin et al., 2018). However, the spatial resolution of satellite data makes it difficult to detect and monitor individual trees that underlie large scale patterns. Individual level data is important for forest ecology, ecosystem services and forestry applications because it connects sets of remote sensing pixels to a fundamental ecological, evolutionary, and economic unit used in analysis. Without grouping to the crown level, it becomes difficult to compare remotely sensed and field-based measurements on individual trees, since field surveys have no corresponding concept of pixels. In addition, characteristics such as species identity, structural traits, growth, and carbon storage potential are properties of individuals rather than pixels. Delineation of crowns also serves as a first step in species classification (Fassnacht et al., 2016), foliar trait mapping (Zheng et al., 2020), and analyses of tree mortality (Stovall et al., 2019).

High-resolution data from airborne sensors have become increasingly accessible, but converting the data into information on individual trees requires significant technical expertise and access to high-performance computing environments (Aubry-Kientz et al., 2019; Puliti et al., 2020). This prevents most ecologists, foresters, and managers from engaging with large scale data on individual trees, despite the availability of the underlying data products and broad importance for forest ecology and management. In response to the growing need for publicly available and standardized airborne remote sensing data over forested ecosystems, the National Ecological Observatory Network (NEON) is collecting multi-sensor data for more than 40 sites across the US. We combine NEON sensor data with a semi-supervised deep learning approach (Weinstein et al., 2019; B. G. Weinstein et al., 2020b) to produce a dataset on the location, height and crown area of over 100 million individual canopy trees at 37 sites distributed

across the United States. To make these data readily accessible, we are releasing easy to access data files to spur biological analyses and to facilitate model development for tasks that rely on individual tree prediction. We describe the components of this open-source dataset, compare predicted crowns with hand-labeled crowns for a range of forest types, and discuss how this dataset can be used in forest research.

The NEON Crowns dataset

The NEON Crowns dataset contains tree crowns for all canopy trees (those visible from airborne remote sensing) at 37 NEON sites. Since subcanopy trees are not visible from above, they are not included in this dataset. We operationally define “trees” as plants over 3m tall. The 37 NEON sites represent all NEON sites containing trees with co-registered RGB and LiDAR data from 2018 or 2019 (see Figure 1 and supplemental table for a list of sites and their locations). Predictions were made using the most recent year for which images were available for each site.

The dataset includes a total of 104,675,304 crowns. Each predicted crown includes data on the spatial position of the crown bounding box, the area of the bounding box (an approximation of crown area), the 99th quantile of the height of LiDAR returns within the bounding box above ground level (an estimate of tree height), the year of sampling, the site where the tree is located, and a confidence score indicating the model confidence that the box represents a tree. The confidence score can vary from 0-1, but based on results from (B. G. Weinstein et al., 2020b), boxes with less than 0.15 confidence were not included in the dataset.

The dataset is provided in two formats: 1) as 11,000 individual files each covering a single 1 km^2 tile (geospatial ‘shapefiles’ in standard ESRI^{TM} format); and 2) as 37 csv files, each covering an entire NEON site. Geospatial tiles have embedded spatial projection information

and can be read in commonly available GIS software (e.g., ArcGIS, QGIS) and geospatial packages for most common programming languages used in data analysis (e.g., R, Python). All data are publicly available, openly licensed (CC-BY), and permanently archived on Zenodo (https://zenodo.org/record/3765872)(B. Weinstein et al., 2020).

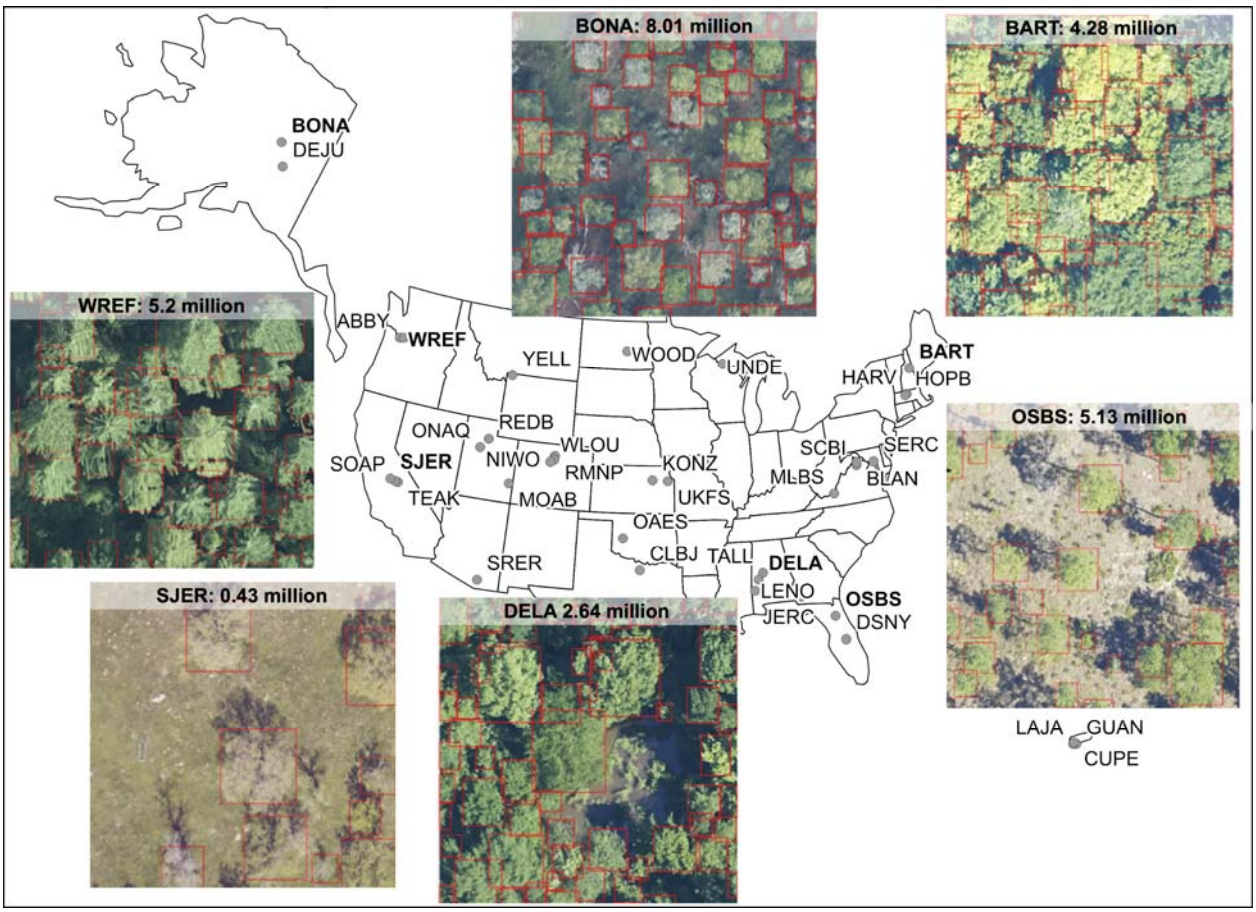


Figure 1. Locations of 37 NEON sites included in the NEON Crowns Dataset and examples of tree predictions shown with RGB imagery for six sites. Clockwise from bottom right: 1) OSBS: Ordway-Swisher Biological Station, Florida 2) DELA: Dead Lake, Alabama, 3) SJER: San Joaquin Experimental Range, California, 4) WREF: Wind River Experimental Forest,

Washington, 5) BONA: Caribou Creek, Alaska and 6) BART: Bartlett Experimental Forest, New Hampshire. Each predicted crown is associated with the spatial position, crown area, maximum height estimates from co-registered LiDAR data, and a predicted confidence score.

To support the visualization of the dataset, we developed a web visualization tool using the ViSUS WebViewer (www.visus.org) to allow users to view all of the trees at the full site scale with the ability to zoom and pan to examine individual groups of trees down to a scale of 20m (see http://visualize.idtrees.org, Figure 2). This tool will allow the ecological community to assist in identifying areas in need of further refinement within the large area covered by the 37 sites.

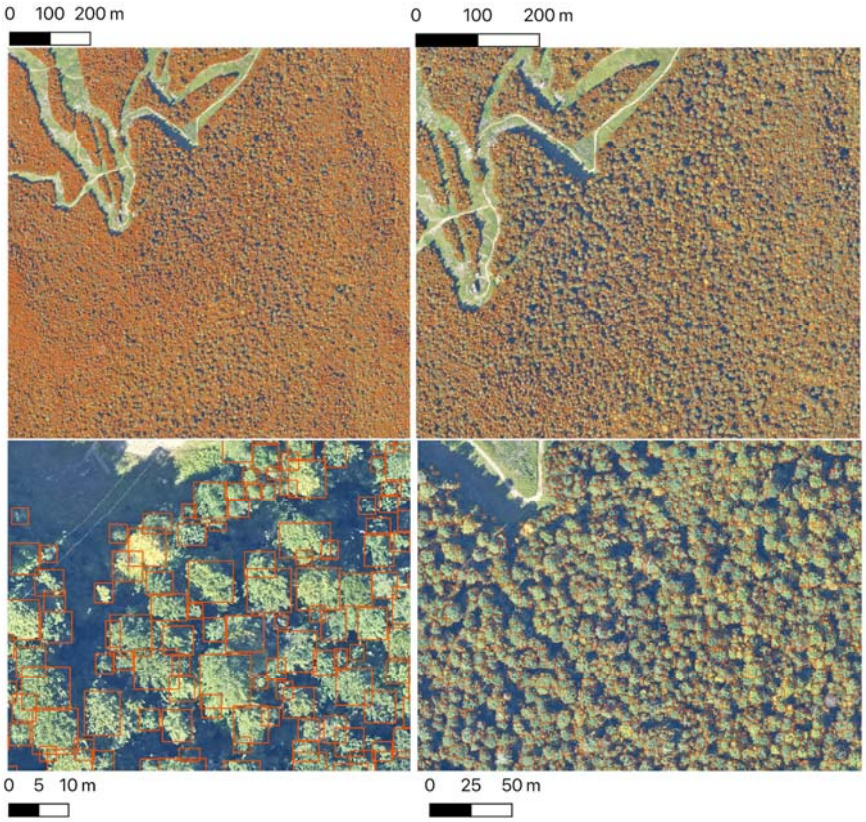


Figure 2. The Neon Crowns Dataset provides individual-level tree predictions at broad scales. An example from Bartlett Forest, NH shows the ability to continuously zoom from landscape

level to stand level views. A single 1km tile is shown. NEON sites tend to have between 100 to 400 tiles in the full airborne footprint.

Crown Delineation

The location of individual tree crowns was estimated using a semi-supervised deep learning workflow (Figure 3) developed in Weinstein et al., (2020b, 2019), which is implemented in the 'DeepForest' Python package (Weinstein et al. 2020a). We extend the workflow by filtering trees using the LiDAR-derived canopy height model to remove objects identified by the model with heights of <3m (Supplementary Material). The deep learning model uses a one-shot object detector with a convolutional neural network backbone to predict tree crowns in RGB imagery. The model was pre-trained first on ImageNet (Deng et al., 2009) and then using weak labels generated from a previous published LiDAR tree detection algorithm using NEON data from 30 sites (Silva et al., 2016). The model was then trained on 10,000 hand-annotated crowns from 7 NEON sites (Figure 1). Hand-annotations included any vegetation over 3m in height, including standing dead trees. The LiDAR derived 3m threshold is important in sparsely vegetated landscapes, such as oak savannah and deserts, where it was difficult for the model to distinguish between trees and low shrubs in the RGB imagery. We chose this approach because it is flexible enough to allow the dataset to be updated and improved by integrating new data and modeling approaches and because it can be effectively applied at large scales with the remote sensing data available from NEON. This required a flexible method that: 1) avoided hand-tuned parameterizations for each site or ecosystem (B. G. Weinstein et al., 2020b), 2) accounted for the highly variable data spanning more than 10,000 tiles that included RGB artifacts and sparse LiDAR point densities, and 3) did not rely on site-specific or species information for allometric constraints on crown size (Duncanson et al., 2015; Fischer et al.,

2020). For details of the underlying algorithms see Weinstein et al. (2019, 2020b). Relaxing each of these constraints opens areas of future improvement, especially once species information is available for each label. For example, Duncanson and Dubayah (2018) showed that site-specific allometric functions can be effective at Teakettle Canyon (TEAK) in predicting tree location and measuring growth over time.

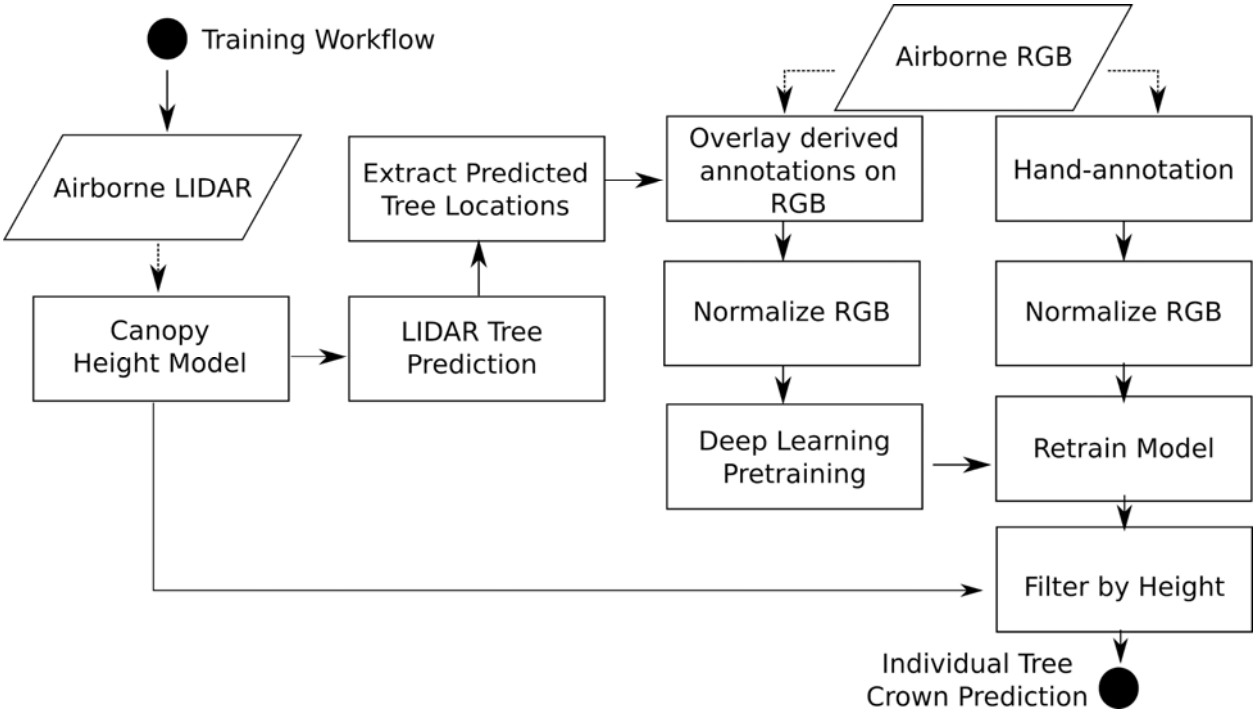


Figure 3. Workflow diagram adapted from (B. G. Weinstein et al., 2020a). The workflow for model training and development are identical to (B. G. Weinstein et al., 2020a) with the exception of extracting heights from the canopy height model for each bounding box prediction.

Evaluation and Validation

The DeepForest method has been compared with leading tree crown detection tools that use an array of sensor data and algorithmic approaches. Weinstein et al., (2020b) compared the approach to three commonly used LiDAR algorithms (Coomes et al., 2017; Li et al., 2012; Silva

et al., 2016) in the lidR package (Roussel et al., 2020) and showed that DeepForest generalized better across forest types with higher precision and recall. Weinstein et al., (2020a) evaluated DeepForest using the data from a recent crown delineation comparison from a tropical forest in French Guiana (Aubry-Kientz et al., 2019). The original paper competed five leading methods (e.g. Ferraz et al., 2016; Hamraz et al., 2016; Williams et al., 2020b) with the authors submitting data to an evaluation dataset kept private by the evaluation team. We repeated this setup and found that DeepForest marginally outperformed all previously tested algorithms, despite the fact that the crown evaluation data used convex polygons and DeepForest used bounding boxes to delineate tree crowns.

In this paper, we further improved the delineation method by incorporating a 3m height filter using the NEON LiDAR-derived canopy model (NEON ID: DP3.30015.001). To validate this addition, we compare predictions to the same set of image-annotated bounding boxes used in Weinstein et al. 2020b (21 NEON sites, 207 images, 6926 trees). Annotations were filtered to 3m in height by comparing bounding boxes. In rare cases, there were obvious trees that were missed by the height threshold. We choose to maintain these rare occurrences as a measure of cross-sensor error when defining ‘tree’ based on an arbitrary lidar-derived height measure. We defined a true positive crown as a predicted bounding box with greater than 50% intersection-over-union (the area of box intersection divided the area of box union of the two boxes) between the predicted and ground truth (image-annotated) bounding box. From the true positives and the total number of samples we calculated crown recall and precision. Crown recall is the proportion of image-annotated crowns matched to a crown prediction and crown precision is the proportion of predictions that match image-annotated crowns. The workflow yielded a bounding box recall of 79.1% with a precision of 72.6%. Tests indicate that the model generalizes well across geographic sites and forest conditions (Figure 4; Weinstein et al., 2020a, 2020b). There is a

general bias towards undersegmenting trees in dense stands where multiple individual trees with similar optical characteristics are grouped into a single delineation. Adding the LiDAR threshold in this implementation resulted in predictions that were 7.0% more precise, but 0.2% less accurate on average (Figure 4). The decrease in recall is due to sparse LiDAR coverage in the CHM model where trees in the evaluation data were clearly taller than 3m were missed in the evaluation dataset.

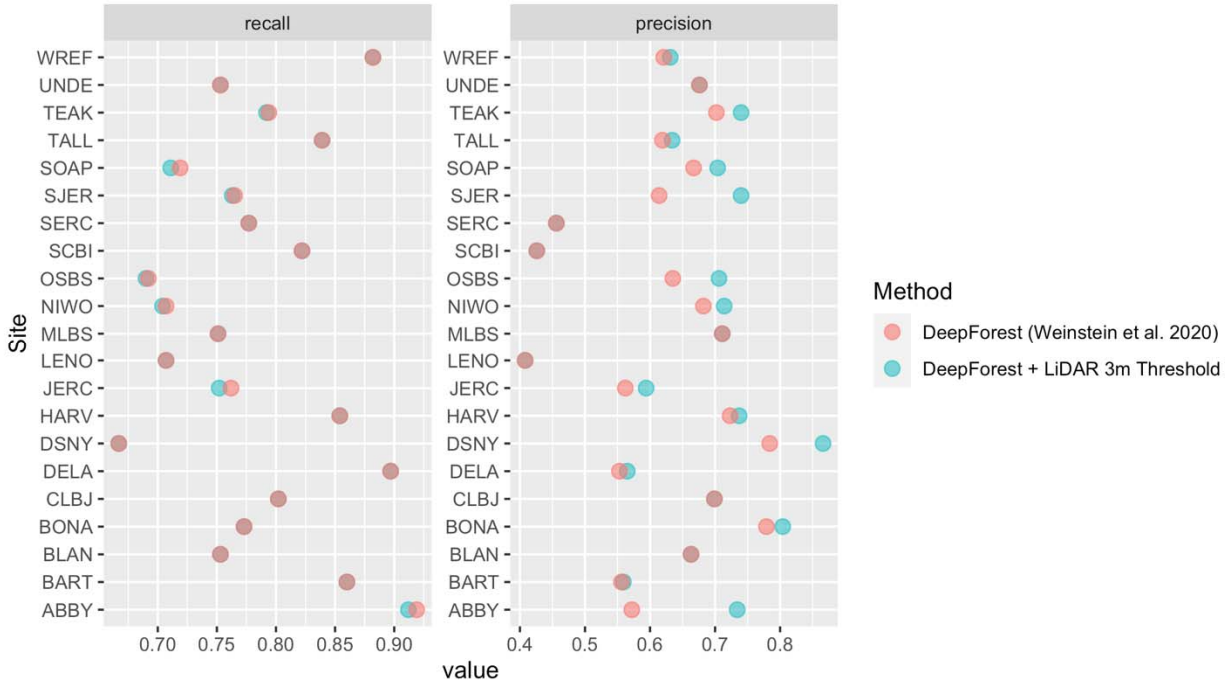


Figure 4. Precision and recall scores for the algorithm used to create the NEON Crowns Dataset (red points), as well as the DeepForest model from Weinstein et al., (2020a) (blue points). Evaluation is performed on 207 image-annotated images (6926 trees) in the NEONTreeEvaluation dataset (https://github.com/weecology/NeonTreeEvaluation). The small drop in recall in the LiDAR thresholding is due to the sparse nature of the LiDAR cloud which can occasionally miss valid trees over 3m.

We also compared crowns delineated by the algorithm to field-collected stems from NEON's Woody Vegetation Structure dataset. This data product contains a single point for each tree with a stem diameter ≥ 10cm. We filtered the raw data to only include live trees likely to be visible in the canopy (see Supplement figure 5 – figure supplement 1). These overstory tree field data help us analyze the performance of our workflow in matching crown predictions to individual trees by scoring the proportion of field stems that fall within a prediction. Field stems can only be applied to one prediction, so if two predictions overlap over a field stem, only one is considered a positive match. This test produces an overall stem recall rate at 69.4%, which is similar to the bounding box recall rate from the image-annotated data (Figure 5). The analysis of stem recall rate is conservative due to the challenge of aligning the field-collected spatial data with the remote sensing data (Figure 5 – figure supplement 1). We found several examples of good predictions that were counted as false positives due to errors in the position of the ground samples within the imagery. The two outliers in OSBS are trees whose most recent field data (2015-2017) are labeled "Live" but have little discernable crowns during leaf-on flights in 2019. It is possible that these trees have since died. In the case of two of the 12 missed trees, they are labeled "disease damaged", and are not recorded in subsequent surveys. Capturing mortality events remains an area of further work, as RGB-based detection requires visible crowns.

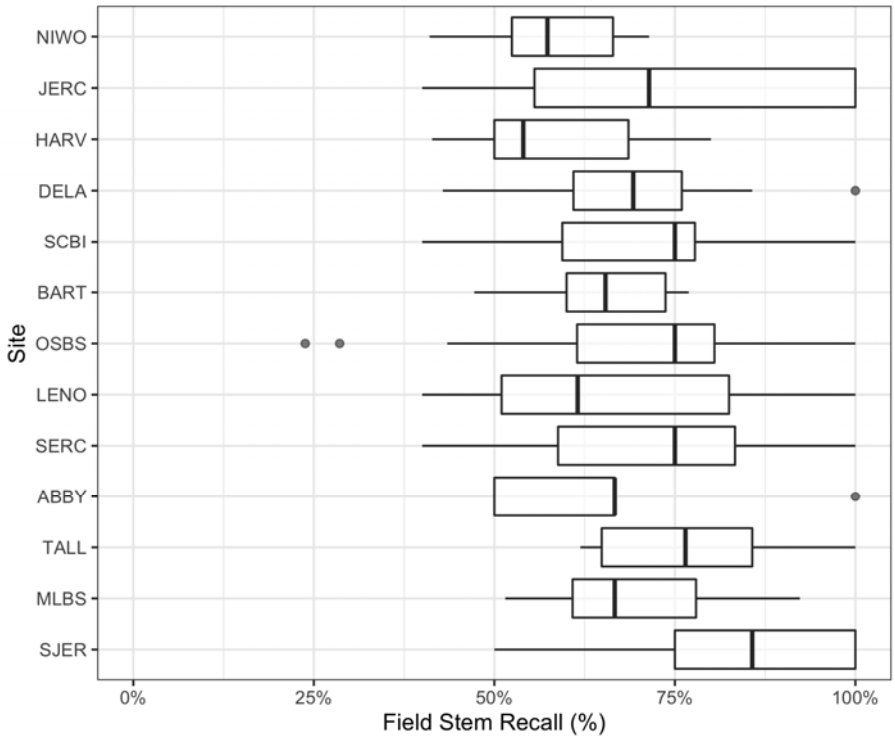


Figure 5. Overstory stem recall rate for NEON sites with available field data. Each data point is the recall rate for a field-collected plot. NEON plots are either 40mx40m 'tower' plots with two 20x20m subplots, or a single 20mx20m 'distributed' plot. See NEON sampling protocols for details. For site abbreviations see Appendix 1 for complete table.

To assess the utility of our approach for mapping forest structure, we compared remotely sensed predictions of maximum tree height to field measurements of tree height of overstory trees using NEON's Woody Plant Vegetation Structure Data. We used the same workflow described in Figure 5 – figure supplement 1 for determining overstory status for both the stem recall and height verification analysis. Predicted heights showed good correspondence with field-measured heights of reference trees. Using a linear-mixed model with a site-level random effect, the predicted crown height had a Root Mean Squared Error of 1.73m (Figure 6). The relationship is stronger in forests with more open canopies (SJER, OSBS) and predictably more

prone to error in forests with denser canopies (BART, MLBS). There is a persistent trend of taller predictions from the remote sensing data as compared with field measured heights. This results in part from tree growth since field measurement due to the temporal gap between field data collection and remote sensing acquisition. For example, 73.8% of the field data for Bartlett Forest (BART; RMSE = 1.68m) came from 2015-2017, but the remote sensing data is from 2019. In addition, previous work to compare field heights to remote sensing data usually first identify trees that are visible from an overhead perspective canopy (e.g. Puliti et al., 2020), whereas all trees above 10cm are sampled in NEON plots. This makes it necessary to infer which crowns are visible (for details for implementation see Figure 5 – figure supplement 1). This process can lead to overestimation of heights if the tree identified in the field data is overtopped by a larger tree, leading to a higher predicted than measured field height. Given the data available, an average RMSE of 1.73m suggests that overstory height measures are reasonably accurate across the dataset.

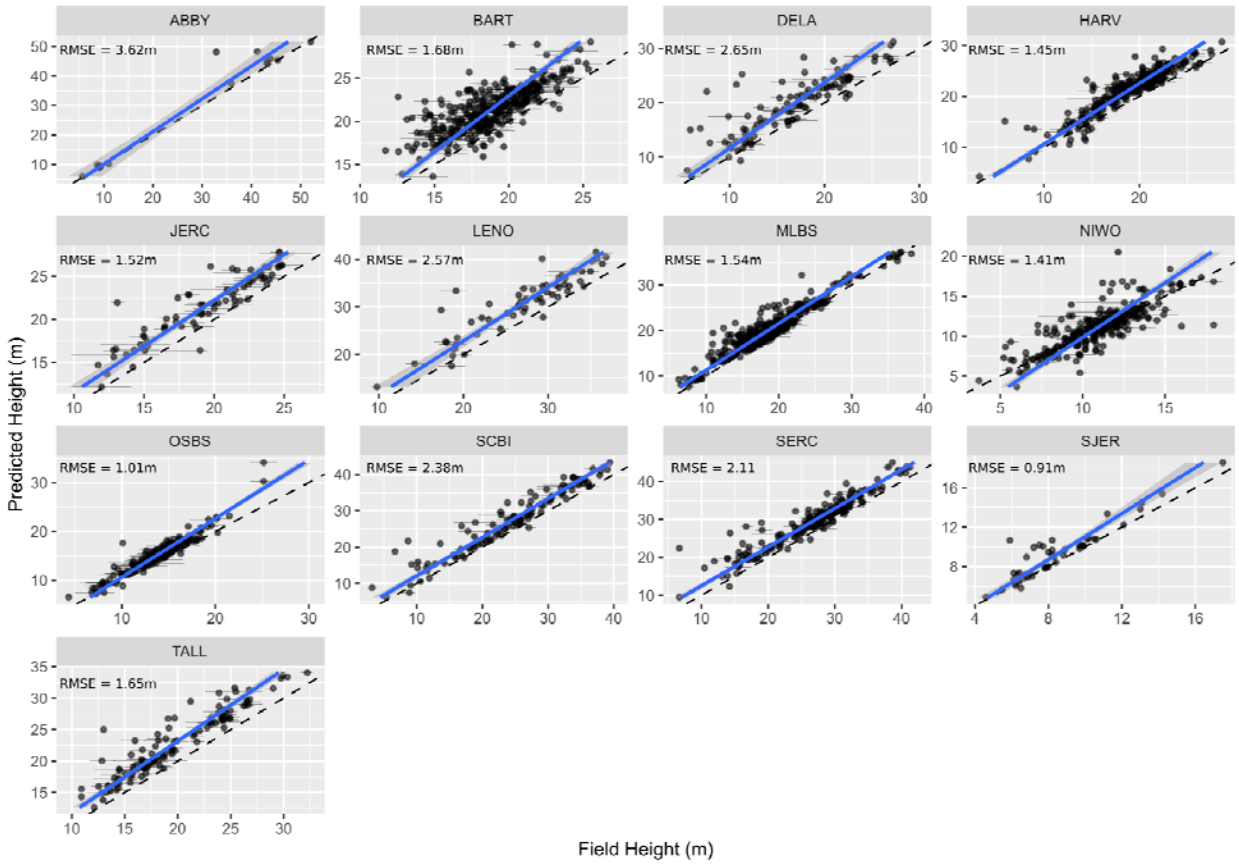


Figure 6. Comparison of field and remote sensing measurements of tree heights for 11 sites in the National Ecological Observatory Network. Each point is an individual tree. See text and Figure 5 – figure supplement 1 for selection criteria and matching scheme for the field data. The RMSE of a mixed-effects model with a site level random effect is 1.73m.

Using the NEON Crowns dataset for individual, landscape and biogeographic scale applications

This dataset supports individual-level cross-scale ecological research that has not been previously possible. It provides the unique combination of information spanning the entire United States, with sites ranging from Puerto Rico to Alaska, with continuous individual-level data within sites at scales hundreds of times larger than what is possible using field-based survey

methods. At the individual level, high-resolution airborne imagery can inform analysis of critical forest properties, such as tree growth and mortality (Clark et al., 2004; Stovall et al., 2019), foliar biochemistry (Chadwick and Asner, 2016; Wang et al., 2020), and landscape-scale carbon storage (S. J. Graves et al., 2018). Because field data on these properties are measured on individual trees, individual level tree detection allows connecting field data directly to image data. In addition, growth, mortality and changes in carbon storage occur on the scale of individual trees such that detection of individual crowns allows direct tracking of these properties across space and time. This allows researchers to understand questions like how individual level attributes relate to mortality in response to disturbance and pests and how the spatial configuration of individual trees within a landscape influences resilience. As a result, this individual level data may be useful for promoting fire resistance landscapes and combating large scale pest outbreaks. While it is possible to aggregate information solely at the stand level, we believe that individual level data opens new possibilities in large scale forest monitoring and provides richer insights into the underlying mechanisms that drive these patterns.

At landscape scales, research is often focused on the effect of environmental and anthropogenic factors on forest structure and biodiversity (Denslow, 1995). For example, understanding why tree biomass and traits vary across landscapes has direct applications to numerous ecological questions and economic implications (e.g. Laubhann et al., 2009). Often this requires sampling at a number of disparate locations and either extrapolation to continuous patterns at landscape scales, or assumptions that the range of possible states of the system are captured by the samples. Using the individual level data from this dataset, we can now produce continuous high-resolution maps across entire NEON sites for enabling landscape scale studies of multiple ecological phenomena (Figure 7). For example, previous work has found that functional and species diversity at local scales promotes biomass and tree growth (Barrufol et

al., 2013; Liang et al., 2016). Similar findings have been reported for phylogenetic diversity at local scales (Satdichanh et al., 2019). Especially when combining with species data, using the crown data to investigate the scale and strength of these effects will inform the mechanisms of community assembly, ecological stability and forest productivity. These landscape scale responses can then be combined with high resolution data on natural and anthropogenic drivers (e.g., topography, soils, fire management) to model forest dynamics at broad scales.

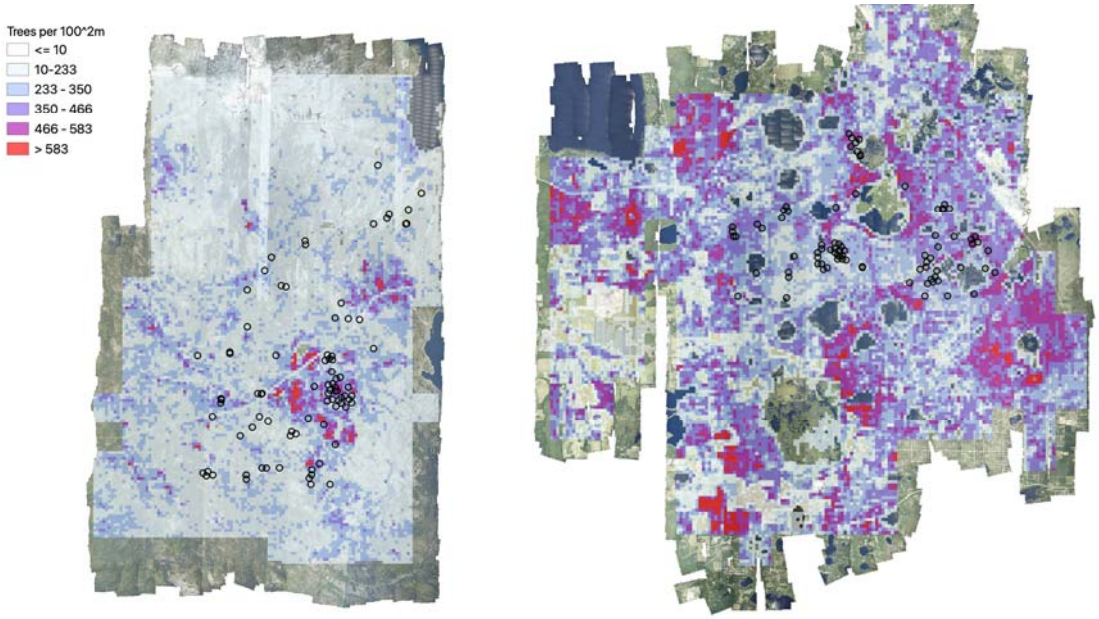


Figure 7. Tree density maps for Teakettle Canyon, California (left) and Ordway Swisher Biological Station, Florida (right). For each 100m^2 pixel, the total number of predicted crowns were counted. The location of NEON Woody Plant Vegetation sampling plots are shown in black circles.

By focusing on detecting individual trees, this approach to landscape scale forest analysis does not require assumptions about spatial similarity, sufficiently extensive sampling, or consistent responses of the ecosystem to drivers across spatial gradients. This is important because the heterogeneity of forest landscapes makes it difficult to use field plot data on

quantities such as tree density and biomass to extrapolate inference to broad scales (Marvin et al., 2014). To illustrate this challenge, we compared field-measured tree densities of NEON field plots to estimated densities of 10,000 remotely sensed plots of the same size placed randomly throughout the landscapes across footprints of the airborne data. We attempted to change the Woody Vegetation data as little as possible (i.e. compared with the more refined filtered data in previous analyses) in order to obtain estimates of tree cover in a plot from the field data. To be included in this analysis, trees needed to have valid spatial coordinates and a minimum height of 3m. Some older data lacked height estimates, in which case we used a minimum DBH threshold of 15cm. In each simulated plot, we then counted the total number of predicted tree crowns to create a distribution of tree densities at the site level (Figure 8). Comparing the field plot tree densities with the distribution from the full site shows deviations for most sites, with NEON field plots exhibiting higher tree densities than encountered on average in the airborne data for some sites (e.g.,Teakettle Canyon, CA) and lower tree densities than from remote sensing in others (e.g., Ordway-Swisher Biological Station). While this kind of comparison is inherently difficult due to differing thresholds and filters for data inclusion in field versus remotely sensed data, it highlights that even well stratified sampling of large landscapes as was done with NEON plots (see NEON technical documents for NEON.DP1.10098) can produce differing tree attribute estimates than continuous sampling from remote sensing data. Combining representative field sampling with remote sensing to produce data products like the NEON Crowns dataset provides an approach to addressing this challenge to improve estimations of the abundance, biomass, and size distributions across large geographic areas.

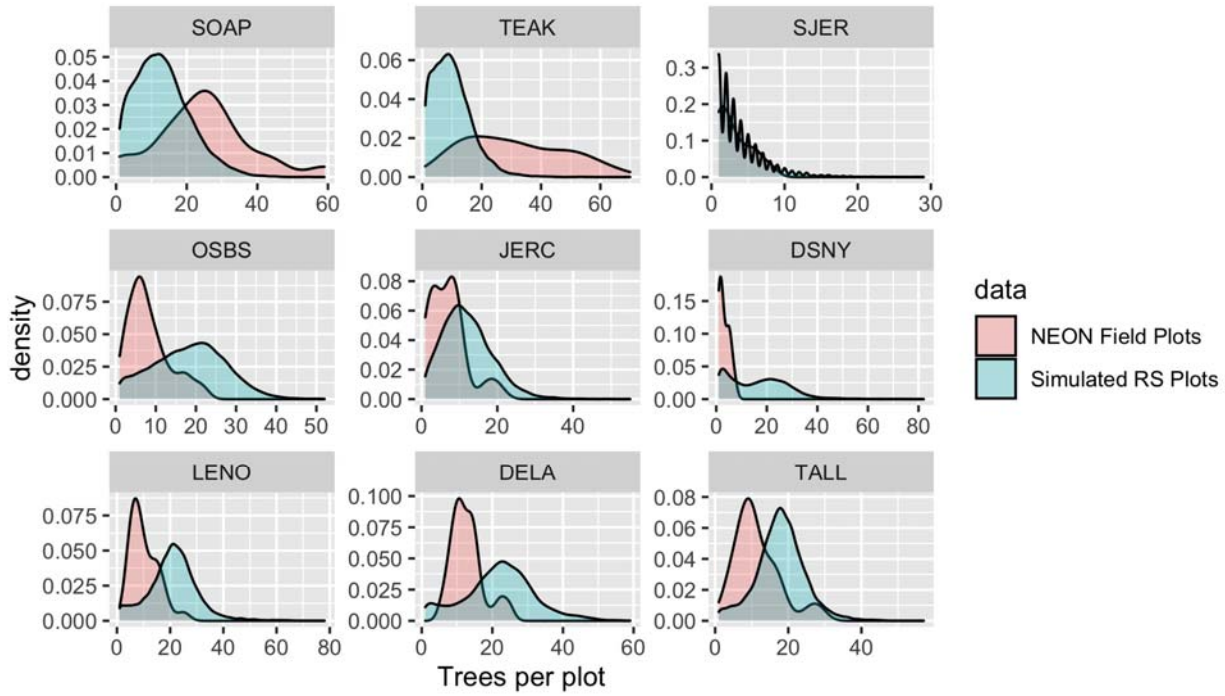


Figure 8. Comparison of tree counts between the field-collected NEON plots and the predicted plots from the dataset. For the remote sensing data, 10000 simulated 40m^2 plots were calculated for each site for the full AOP footprint for each year. To mimic NEON sampling, 2 quadrants were randomly sampled in each simulated plot. No plots on water, bare ground, or herbaceous land classes were included in the comparison. We selected three sites from three NEON domains to show a sample of sites across the continental US. Both distributed and tower NEON plots were used for these analyses.

The NEON Crowns dataset supports the assessment of cross-site patterns to help understand the influence of large-scale processes on forest structure at biogeographic scales. For example, ecologists are interested in how and why forest characteristics such as abundance, biomass, and allometric relationships vary among forest types (e.g. Jucker et al., 2017) and how these patterns covary across environmental gradients (Feldpausch et al., 2011). Understanding these relationships is important for inferring controls over forest stand structure,

understanding individual tree biology, and assessing stand productivity. For example, are local patterns of density and structural biomass primarily the result of historical mechanisms, such as dispersal and adaptation, or local mechanisms such as nutrient availability? By providing standardized data that span near-continental scales, this dataset can help inform the fundamental mechanisms that shape forest structure and dynamics. For example, we can calculate tree allometries (e.g., the ratio of tree height to crown area) on a large number of individual trees across NEON sites and explore how allometry varies with stand density and vegetation type (Figure 9). This analysis shows a continental-scale relationship, with denser forests exhibiting trees with narrower crowns for the same tree height compared with less dense forests, but also clustering and variation in the relationship within vegetation types. For example, subalpine forests illustrate relationships between tree density and allometry that are distinct from other forest types. By defining both general biogeographic patterns, and deviations therein, this dataset will allow the investigation of factors shaping forest structure at macroecological scales.

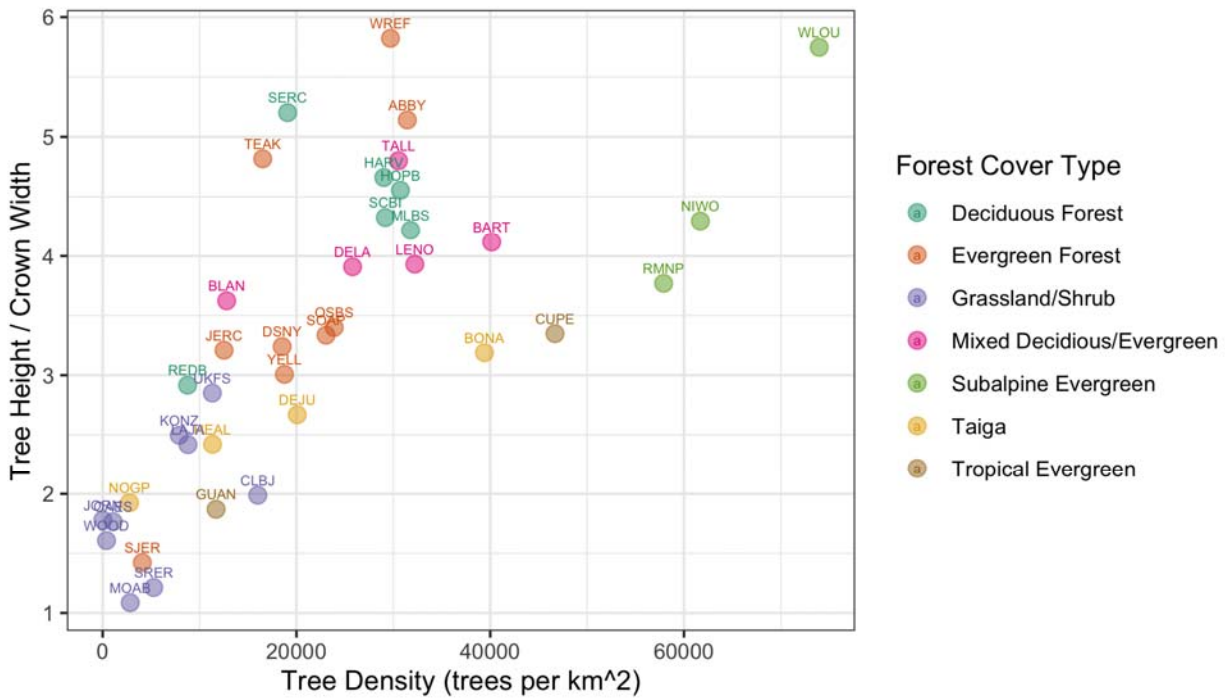


Figure 9. Individual crown attributes for predictions made at each NEON site. For site abbreviations see Appendix 1. Crown area is calculated by multiplying the width and height of the predicted crown bounding box. Crown height is the 99th quantile of the LiDAR returns that fall inside the predicted crown bounding box. Sites are colored by the dominant forest type to illustrate the general macroecological relationship among sites in similar biomes.

In addition to these ecological applications, the NEON Crowns dataset can also act as a foundation for other machine learning and computer vision applications in forest informatics, such as tree health assessments, species classification, or foliar trait estimation both within NEON sites (Wang et al., 2020) and outside of NEON sites (Schneider et al., 2020). In each of these tasks, individual tree delineation is the first step to associate sensor data with ground measurements. However, delineation requires a distinct set of technical background and computational approaches and thus many ecological applications skip an explicit delineation step entirely (Williams et al., 2020a). In addition, the growing availability of continental scale

datasets of high resolution remote sensing imagery opens up the possibility for broad scale forest monitoring of individual trees (Brandt et al., 2020; Schneider et al., 2020) that can be supported by this dataset. Just as we used weak annotations generated from unsupervised LiDAR algorithms, future developers can use this dataset to train in the multiple data types provided by the NEON Airborne Observation Platform across a broad range of forest types. While our crown annotations are not perfect, they are specifically tailored to one of the largest datasets that allows pairing individual tree detections with information on species identity, tree health, and leaf traits through NEONs field sampling, and we believe they are sufficiently robust to serve as the basis for broad scale analysis.

Limitations and Further Technical Developments

An important limitation for this dataset is that it only provides information on sun-exposed tree crowns. It is therefore not appropriate for ecological analyses that depend on accurate characterization of subcanopy trees and the three-dimensional structure of forest stands. Fortunately, a number of the major questions and applications in ecology are primarily influenced by large individuals (Enquist et al., 2020). For example, biomass estimation is largely driven by the canopy in most ecosystems, rather than mid or understory trees that are likely to be missed by aerial surveys. Similarly, habitat classification and species abundance curves can depend on the dominant forest structure that can be inferred from coarse resolution airborne data (Shirley et al., 2013) and could be improved using this dataset. It may be possible to establish relationships between understory and canopy measures using field data that could allow this dataset to be used as part of a broader analysis (Bohlman, 2015; Duncanson et al., 2014; Fischer et al., 2020). However, this would require significant additional research to validate the potential for this type of approach at continental scales.

We experimented with avenues to combine RGB information and a LiDAR canopy height model to create a jointly learned input and found that no combination of data fusion outperformed the current pipeline (Figure 4 – figure supplement 2). The lack of improvement when directly incorporating LiDAR data into the CNN is likely due to a combination of geographic variation in tree shape and LiDAR coverage, sparse LiDAR point densities (~6 pts/m at many NEON sites), and a lack of joint RGB and LiDAR data for initial pretraining. Most LiDAR based methods are evaluated on data from a single forest type with point densities ranging from ~15 pts/m (e.g., Duncanson et al. 2018) to over 100 pts/m (e.g., Aubry-Kientz et al. 2019). As instrumentation improves to support collecting higher density LiDAR consistently at larger scales and algorithms are improved to allow generalization across forest types, we anticipate updating the dataset with improved delineation of the sunlit canopy and begin to add subcanopy trees.

An additional limitation is the uncertainty inherent in the algorithmic detection of crowns. While we found good correspondence between image-based crown annotations and those produced by the model for many sites, there remained substantial uncertainty across all sites and reasonable levels of error in some sites. It is important to consider how this uncertainty will influence the inference from research using this and similar datasets. The model is biased towards undersegmentation, meaning that multiple trees are prone to being grouped as a single crown. It is also somewhat conservative in estimating crown extent wherein it tends to ignore small extensions of branches from the main crown. These biases could impact studies of tree allometry and biomass if the analysis is particularly sensitive to crown area. When making predictions for ecosystem features such as biomass, it will be important to propagate the uncertainty in individual crowns into downstream analyses. While confidence scores for

individual detections are provided to aid uncertainty propagation, the use of additional ground truth data may also be necessary to infer reliability.

One aspect of individual crown uncertainty that we have not addressed is the uncertainty related to image-based crown annotations and measurement of trees in the field (S. Graves et al., 2018). To allow training and evaluating the model across a broad range of forest types, we used image-based crown annotations. This approach assumes that crowns identifiable in remotely sensed imagery accurately reflect trees on the ground. This will not always be the case, as what appears to be a single crown from above may constitute multiple neighboring trees, and conversely, what appears to be two distinct crowns in an image may be two branches of a single large tree (S. Graves et al., 2018). Distinguishing individual trees, especially when considering species with multi-bole stems can be subjective, even during field surveys. Targeted field surveys will be always needed to validate these predictions and community annotation efforts will allow for assessment of this component of uncertainty. In particular, combining terrestrial LiDAR sampling with airborne sensors is a promising route to both validate the number of stems and establish subcanopy diversity (Calders et al., 2020). In addition, when co-registered hyperspectral data are available, it may help to separate neighboring trees in diverse forests, provided it does not cause lumping of neighboring trees of the same species. Weighing these tradeoffs across a range of forest types remains an open area of exploration.

The machine learning workflow used to generate this dataset also has several areas that could be improved for greater accuracy, transferability and robustness. The current model contains a single class ‘Tree’ with an associated confidence score. This representation prevents the model from differentiating between objects that are not trees and objects for which sufficient training information is not available. For example, the model has been trained to ignore buildings and other vertical structures that may look like trees. However, when confronted by

objects data that has never been encountered, it often produces unintuitive results. Examples of objects that did not appear in the training data, and as a result were erroneously predicted as trees, include weather stations, floating buoys, and oil wells. Designing models that can identify outliers, anomalies, and ‘unknown’ objects is an active area of research in machine learning and will be useful in increasing accuracy in novel environments. In addition, NEON data can sometimes be afflicted by imaging artifacts due to co-registration issues with LiDAR and raw RGB imagery (Figure 7 – figure supplement 1). This effect can lead to distorted imagery that appears fuzzy and swirled and lead to poor segmentation. An ideal model would detect these areas of poor quality and label them as ‘unknown’ rather than attempting to detect trees in these regions.

Given these limitations, we view this version of the dataset as the first step in an iterative process to improve cross-scale individual level data on trees. Ongoing assessment of these predictions using both our visualization tool and field-based surveys will be crucial to continually identify areas for improvements in both training data and modeling approaches. While iterative improvements are important, the accuracy of the current predictions illustrates that this dataset is sufficiently precise for addressing many cross-scale questions related to forest structure. By providing broad scale crown data we hope to highlight the promising integration between deep learning, remote sensing, and forest informatics, and provide access to the results of this next key step in ecological research to the broad range of stakeholders who can benefit from these data.

Acknowledgements

We would like to thank NEON staff and in particular Tristan Goulden and Courtney Meier for their assistance and support. This research was supported by the Gordon and Betty Moore

Foundation's Data-Driven Discovery Initiative (GBMF4563) to E.P. White and by the National Science Foundation (1926542) to E.P. White, S.A. Bohlman, A. Zare, D.Z. Wang, and A. Singh and USDA National Institute of Food and Agriculture McIntire Stennis projects #1007080 to S. A. Bohlman. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Literature Cited

Aubry-Kientz M, Dutrieux R, Ferraz A, Saatchi S, Hamraz H, Williams J, Coomes D, Piboule A, Vincent G. 2019. A Comparative Assessment of the Performance of Individual Tree Crowns Delineation Algorithms from ALS Data in Tropical Forests. *Remote Sens* **11**:1086. doi:10.3390/rs11091086

Barrufol M, Schmid B, Bruelheide H, Chi X, Hector A, Ma K, Michalski S, Tang Z, Niklaus PA. 2013. Biodiversity Promotes Tree Growth during Succession in Subtropical Forest. *PLoS ONE* **8**:e81246. doi:10.1371/journal.pone.0081246

Bastin J-F, Rutishauser E, Kellner JR, Saatchi S, Pélissier R, Hérault B, Slik F, Bogaert J, De Cannière C, Marshall AR, Poulsen J, Alvarez-Loyayza P, Andrade A, Angbonga-Basia A, Araujo-Murakami A, Arroyo L, Ayyappan N, de Azevedo CP, Banki O, Barbier N, Barroso JG, Beeckman H, Bitariho R, Boeckx P, Boehning-Gaese K, Brandão H, Brearley FQ, Breuer Ndoundou Hockemba M, Brienen R, Camargo JLC, Campos-Arceiz A, Cassart B, Chave J, Chazdon R, Chuyong G, Clark DB, Clark CJ, Condit R, Honorio Coronado EN, Davidar P, de Haulleville T, Descroix L, Doucet J-L, Dourdain A, Droissart V, Duncan T, Silva Espejo J, Espinosa S, Farwig N, Fayolle A, Feldpausch TR, Ferraz A, Fletcher C, Gajapersad K, Gillet J-F, Amaral IL do, Gonmadje C, Grogan J, Harris D, Herzog SK, Homeier J, Hubau W, Hubbell SP, Hufkens K, Hurtado J, Kamdem NG, Kearsley E, Kenfack D, Kessler M, Labrière N, Laumonier Y, Laurance S, Laurance WF, Lewis SL, Libalah MB, Ligot G, Lloyd J, Lovejoy TE, Malhi Y, Marimon BS, Marimon Junior BH, Martin EH, Matius P, Meyer V, Mendoza Bautista C, Monteagudo-Mendoza A, Mtui A, Neill D, Parada Gutierrez GA, Pardo G, Parren M, Parthasarathy N, Phillips OL, Pitman NCA, Ploton P, Ponette Q, Ramesh BR, Razafimahaimodison J-C, Réjou-Méchain M, Rolim SG, Saltos HR, Rossi LMB, Spironello WR, Rovero F, Saner P, Sasaki D, Schulze M, Silveira M, Singh J, Sist P, Sonke B, Soto JD, de Souza CR, Stropp J, Sullivan MJP, Swanepoel B, Steege H ter, Terborgh J, Texier N, Toma T, Valencia R, Valenzuela L, Ferreira LV, Valverde FC, Van Andel TR, Vasque R, Verbeeck H, Vivek P, Vleminckx J, Vos VA, Wagner FH, Warsudi PP, Wortel V, Zagt RJ, Zebaze D. 2018. Pan-tropical prediction of forest structure from the largest trees. *Glob Ecol Biogeogr* **27**:1366–1383. doi:10.1111/geb.12803

Bohlman SA. 2015. Species Diversity of Canopy Versus Understory Trees in a Neotropical Forest: Implications for Forest Structure, Function and Monitoring. *Ecosystems* **18**:658–670. doi:10.1007/s10021-015-9854-0

Brandt M, Tucker CJ, Kariryaa A, Rasmussen K, Abel C, Small J, Chave J, Rasmussen LV, Hiernaux P, Diouf AA, Kergoat L, Mertz O, Igel C, Gieseke F, Schöning J, Li S, Melocik

K, Meyer J, Sinno S, Romero E, Glennie E, Montagu A, Dendoncker M, Fensholt R. 2020. An unexpectedly large count of trees in the West African Sahara and Sahel. *Nature* 1–5. doi:10.1038/s41586-020-2824-5

Calders K, Adams J, Armston J, Bartholomeus H, Bauwens S, Bentley LP, Chave J, Danson FM, Demol M, Disney M, Gaulton R, Krishna Moorthy SM, Levick SR, Saarinen N, Schaaf C, Stovall A, Terryn L, Wilkes P, Verbeeck H. 2020. Terrestrial laser scanning in forest ecology: Expanding the horizon. *Remote Sens Environ* **251**:112102. doi:10.1016/j.rse.2020.112102

Chadwick KD, Asner GP. 2016. Organismic-Scale Remote Sensing of Canopy Foliar Traits in Lowland Tropical Forests. *Remote Sens* **8**:87. doi:10.3390/rs8020087

Clark DB, Castro CS, Alvarado LDA, Read JM. 2004. Quantifying mortality of tropical rain forest trees using high-spatial-resolution satellite data. *Ecol Lett* **7**:52–59. doi:10.1046/j.1461-0248.2003.00547.x

Coomes DA, Dalponte M, Jucker T, Asner GP, Banin LF, Burslem DFRP, Lewis SL, Nilus R, Phillips OL, Phua M-H, Qie L. 2017. Area-based vs tree-centric approaches to mapping forest carbon in Southeast Asian forests from airborne laser scanning data. *Remote Sens Environ* **194**:77–88. doi:10.1016/j.rse.2017.03.017

Crowther TW, Glick HB, Covey KR, Bettigole C, Maynard DS, Thomas SM, Smith JR, Hintler G, Duguid MC, Amatulli G, Tuanmu M-N, Jetz W, Salas C, Stam C, Piotto D, Tavani R, Green S, Bruce G, Williams SJ, Wiser SK, Huber MO, Hengeveld GM, Nabuurs G-J, Tikhonova E, Borchardt P, Li C-F, Powrie LW, Fischer M, Hemp A, Homeier J, Cho P, Vibrans AC, Umunay PM, Piao SL, Rowe CW, Ashton MS, Crane PR, Bradford MA. 2015. Mapping tree density at a global scale. *Nature* **525**:201–205. doi:10.1038/nature14967

Deng J, Dong W, Socher R, Li L, Kai Li, Li Fei-Fei. 2009. ImageNet: A large-scale hierarchical image database2009 IEEE Conference on Computer Vision and Pattern Recognition. Presented at the 2009 IEEE Conference on Computer Vision and Pattern Recognition. pp. 248–255. doi:10.1109/CVPR.2009.5206848

Denslow JS. 1995. Disturbance and Diversity in Tropical Rain Forests: The Density Effect. *Ecol Appl* **5**:962–968. doi:https://doi.org/10.2307/2269347

Duncanson L, Dubayah R. 2018. Monitoring individual tree-based change with airborne lidar. *Ecol Evol* **8**:5079–5089. doi:10.1002/ece3.4075

Duncanson L, Rourke O, Dubayah R. 2015. Small Sample Sizes Yield Biased Allometric Equations in Temperate Forests. *Sci Rep* **5**:17153. doi:10.1038/srep17153

Duncanson LI, Cook BD, Hurtt GC, Dubayah RO. 2014. An efficient, multi-layered crown delineation algorithm for mapping individual tree structure across multiple ecosystems. *Remote Sens Environ* **154**:378–386. doi:10.1016/j.rse.2013.07.044

Enquist BJ, Link to external site this link will open in a new window, Abraham AJ, Link to external site this link will open in a new window, J HMB, Yadvinder M, Link to external site this link will open in a new window, Doughty CE, Link to external site this link will open in a new window. 2020. The megabiota are disproportionately important for biosphere functioning. *Nat Commun Lond* **11**. doi:http://dx.doi.org.lp.hscl.ufl.edu/10.1038/s41467-020-14369-y

Fassnacht FE, Latifi H, Stereńczak K, Modzelewska A, Lefsky M, Waser LT, Straub C, Ghosh A. 2016. Review of studies on tree species classification from remotely sensed data. *Remote Sens Environ* **186**:64–87. doi:10.1016/j.rse.2016.08.013

Feldpausch TR, Banin L, Phillips OL, Baker TR, Lewis SL, Quesada CA, Affum-Baffoe K, Arets EJMM, Berry NJ, Bird M, Brondizio ES, de Camargo P, Chave J, Djagbletey G,

Domingues TF, Drescher M, Fearnside PM, França MB, Fyllas NM, Lopez-Gonzalez G, Hladik A, Higuchi N, Hunter MO, Iida Y, Salim KA, Kassim AR, Keller M, Kemp J, King DA, Lovett JC, Marimon BS, Marimon-Junior BH, Lenza E, Marshall AR, Metcalfe DJ, Mitchard ETA, Moran EF, Nelson BW, Nilus R, Nogueira EM, Palace M, Patiño S, Peh KS-H, Raventos MT, Reitsma JM, Saiz G, Schrodt F, Sonké B, Taedoumg HE, Tan S, White L, Wöll H, Lloyd J. 2011. Height-diameter allometry of tropical forest trees. *Biogeosciences* **8**:1081–1106. doi:10.5194/bg-8-1081-2011

Ferraz A, Saatchi S, Mallet C, Meyer V. 2016. Lidar detection of individual tree size in tropical forests. *Remote Sens Environ* **183**:318–333. doi:10.1016/j.rse.2016.05.028

Fischer FJ, Labrière N, Vincent G, Hérault B, Alonso A, Memiaghe H, Bissiengou P, Kenfack D, Saatchi S, Chave J. 2020. A simulation method to infer tree allometry and forest structure from airborne laser scanning and forest inventories. *Remote Sens Environ* **251**:112056. doi:10.1016/j.rse.2020.112056

Graves S, Gearhart J, Caughlin TT, Bohlman S. 2018. A digital mapping method for linking high-resolution remote sensing images to individual tree crowns (preprint). PeerJ Preprints. doi:10.7287/peerj.preprints.27182v1

Graves SJ, Caughlin TT, Asner GP, Bohlman SA. 2018. A tree-based approach to biomass estimation from remote sensing data in a tropical agricultural landscape. *Remote Sens Environ* **218**:32–43. doi:10.1016/j.rse.2018.09.009

Hamraz H, Contreras MA, Zhang J. 2016. A robust approach for tree segmentation in deciduous forests using small-footprint airborne LiDAR data. *Int J Appl Earth Obs Geoinformation* **52**:532–541. doi:10.1016/j.jag.2016.07.006

Hansen MC, Potapov PV, Moore R, Hancher M, Turubanova SA, Tyukavina A, Thau D, Stehman SV, Goetz SJ, Loveland TR, Kommareddy A, Egorov A, Chini L, Justice CO, Townshend JRG. 2013. High-Resolution Global Maps of 21st-Century Forest Cover Change. *Science* **342**:850–853. doi:10.1126/science.1244693

Jucker T, Caspersen J, Chave J, Antin C, Barbier N, Bongers F, Dalponte M, Ewijk KY van, Forrester DI, Haeni M, Higgins SI, Holdaway RJ, Iida Y, Lorimer C, Marshall PL, Momo S, Moncrieff GR, Ploton P, Poorter L, Rahman KA, Schlund M, Sonké B, Sterck FJ, Trugman AT, Usoltsev VA, Vanderwel MC, Waldner P, Wedeux BMM, Wirth C, Wöll H, Woods M, Xiang W, Zimmermann NE, Coomes DA. 2017. Allometric equations for integrating remote sensing imagery into forest monitoring programmes. *Glob Change Biol* **23**:177–190. doi:10.1111/gcb.13388

Laubhann D, Sterba H, Reinds GJ, De Vries W. 2009. The impact of atmospheric deposition and climate on forest growth in European monitoring plots: An individual tree growth model. *For Ecol Manag* **258**:1751–1761. doi:10.1016/j.foreco.2008.09.050

Li W, Guo Q, Jakubowski MK, Kelly M. 2012. A New Method for Segmenting Individual Trees from the Lidar Point Cloud. *Photogramm Eng Remote Sens* **78**:75–84. doi:10.14358/PERS.78.1.75

Liang J, Crowther TW, Picard N, Wiser S, Zhou M, Alberti G, Schulze E-D, McGuire AD, Bozzato F, Pretzsch H, de-Miguel S, Paquette A, Hérault B, Scherer-Lorenzen M, Barrett CB, Glick HB, Hengeveld GM, Nabuurs G-J, Pfautsch S, Viana H, Vibrans AC, Ammer C, Schall P, Verbyla D, Tchebakova N, Fischer M, Watson JV, Chen HYH, Lei X, Schelhaas M-J, Lu H, Gianelle D, Parfenova EI, Salas C, Lee E, Lee B, Kim HS, Bruelheide H, Coomes DA, Piotto D, Sunderland T, Schmid B, Gourlet-Fleury S, Sonké B, Tavani R, Zhu J, Brandl S, Vayreda J, Kitahara F, Searle EB, Neldner VJ, Ngugi MR, Baraloto C, Frizzera L, Bałazy R, Oleksyn J, Zawiła-Niedźwiecki T, Bouriaud O, Bussotti F, Finér L, Jaroszewicz B, Jucker T, Valladares F, Jagodzinski AM, Peri PL, Gonmadje

C, Marthy W, O'Brien T, Martin EH, Marshall AR, Rovero F, Bitariho R, Niklaus PA, Alvarez-Loayza P, Chamuya N, Valencia R, Mortier F, Wortel V, Engone-Obiang NL, Ferreira LV, Odeke DE, Vasquez RM, Lewis SL, Reich PB. 2016. Positive biodiversity-productivity relationship predominant in global forests. *Science* **354**. doi:10.1126/science.aaf8957

Marvin DC, Asner GP, Knapp DE, Anderson CB, Martin RE, Sinca F, Tupayachi R. 2014. Amazonian landscapes and the bias in field studies of forest structure and biomass. *Proc Natl Acad Sci* **111**:E5224–E5232. doi:10.1073/pnas.1412999111

Maschler J, Atzberger C, Immitzer M. 2018. Individual Tree Crown Segmentation and Classification of 13 Tree Species Using Airborne Hyperspectral Data. *Remote Sens* **10**:1218. doi:10.3390/rs10081218

Puliti S, Breidenbach J, Astrup R. 2020. Estimation of Forest Growing Stock Volume with UAV Laser Scanning Data: Can It Be Done without Field Data? *Remote Sens* **12**:1245. doi:10.3390/rs12081245

Roussel J-R, Auty D, Coops NC, Tompalski P, Goodbody TRH, Meador AS, Bourdon J-F, de Boissieu F, Achim A. 2020. lidR: An R package for analysis of Airborne Laser Scanning (ALS) data. *Remote Sens Environ* **251**:112061. doi:10.1016/j.rse.2020.112061

Satdichanh M, Ma H, Yan K, Dossa GGO, Winowiecki L, Vågen T-G, Gassner A, Xu J, Harrison RD. 2019. Phylogenetic diversity correlated with above-ground biomass production during forest succession: Evidence from tropical forests in Southeast Asia. *J Ecol* **107**:1419–1432. doi:https://doi.org/10.1111/1365-2745.13112

Schneider FD, Ferraz AA, Hancock S, Duncanson LI, Dubayah RO, Pavlick RP, Schimel DS. 2020. Towards mapping the diversity of canopy structure from space with GEDI. *Environ Res Lett*. doi:10.1088/1748-9326/ab9e99

Shirley SM, Yang Z, Hutchinson RA, Alexander JD, McGarigal K, Betts MG. 2013. Species distribution modelling for the people: unclassified landsat TM imagery predicts bird occurrence at fine resolutions. *Divers Distrib* **19**:855–866. doi:10.1111/ddi.12093

Silva CA, Hudak AT, Vierling LA, Loudermilk EL, O'Brien JJ, Hiers JK, Jack SB, Gonzalez-Benecke C, Lee H, Falkowski MJ, Khosravipour A. 2016. Imputation of Individual Longleaf Pine (*Pinus palustris* Mill.) Tree Attributes from Field and LiDAR Data. *Can J Remote Sens* **42**:554–573. doi:10.1080/07038992.2016.1196582

Stovall AEL, Shugart H, Yang X. 2019. Tree height explains mortality risk during an intense drought. *Nat Commun* **10**:4385. doi:10.1038/s41467-019-12380-6

Wang Z, Chlus A, Geygan R, Ye Z, Zheng T, Singh A, Couture JJ, Cavender-Bares J, Kruger EL, Townsend PA. 2020. Foliar functional traits from imaging spectroscopy across biomes in eastern North America. *New Phytol* **228**:494–511. doi:10.1111/nph.16711

Weinstein B, Marconi S, Zare, A A, Bohlman S, Graves S, Singh A, White E. 2020. NEON Tree Crowns Dataset. doi:10.5281/zenodo.3765872

Weinstein BG, Marconi S, Aubry-Kientz, Mélaine M, Vincent G, Senyondo H, White E. 2020a. DeepForest: A Python package for RGB deep learning tree crown delineation. *Methods Ecol Evol*.

Weinstein BG, Marconi S, Bohlman S, Zare A, White E. 2019. Individual Tree-Crown Detection in RGB Imagery Using Semi-Supervised Deep Learning Neural Networks. *Remote Sens* **11**:1309. doi:10.3390/rs11111309

Weinstein BG, Marconi S, Bohlman SA, Zare A, White EP. 2020b. Cross-site learning in deep learning RGB tree crown detection. *Ecol Inform* **56**:101061. doi:10.1016/j.ecoinf.2020.101061

Williams J, Schönlieb C-B, Swinfield T, Irawan B, Achmad E, Zudhi M, Habibi, Gemita E, Coomes DA. 2020a. SLIC-UAV: A Method for monitoring recovery in tropical restoration projects through identification of signature species using UAVs. *ArXiv200606624 Cs Stat*.

Williams J, Schönlieb C-B, Swinfield T, Lee J, Cai X, Qie L, Coomes DA. 2020b. Three-dimensional Segmentation of Trees Through a Flexible Multi-Class Graph Cut Algorithm (MCGC). *IEEE Trans Geosci Remote Sens* **58**:754–776. doi:10.1109/TGRS.2019.2940146

Zheng Z, Zeng Y, Schneider FD, Zhao Y, Zhao D, Schmid B, Schaepman ME, Morsdorf F. 2020. Mapping functional diversity using individual tree-based morphological and physiological traits in a subtropical forest. *Remote Sens Environ* 112170. doi:10.1016/j.rse.2020.112170

Supplemental Figures

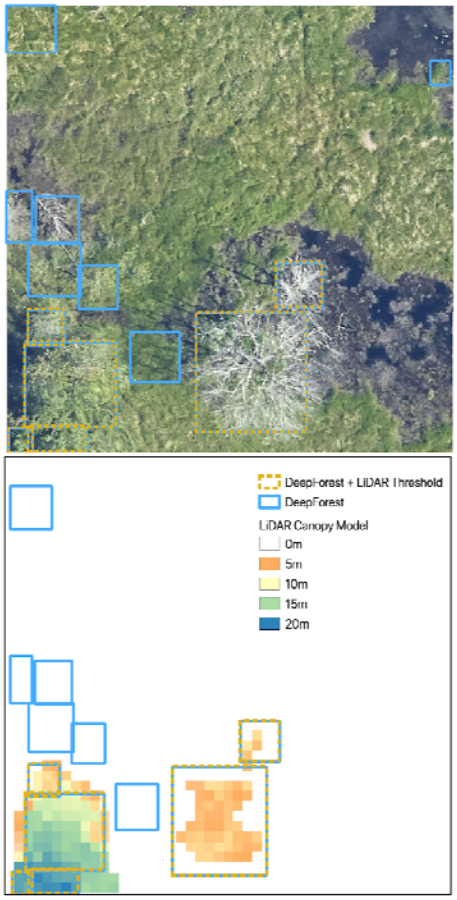
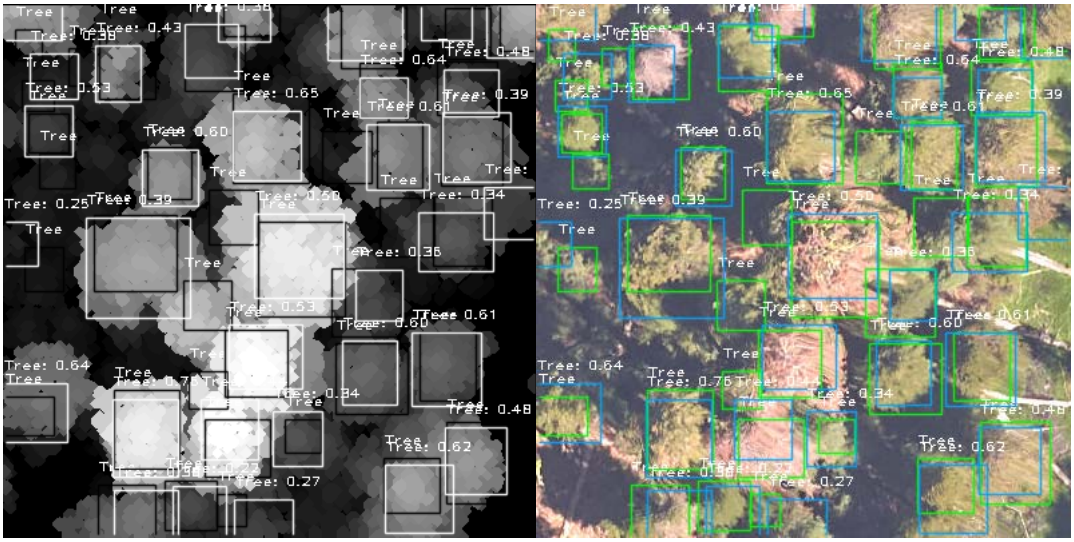


Figure 4 - figure supplement 1. Illustration of the LiDAR threshold for minimum predicted tree height. To utilize LiDAR information in the post-processing step we decided on a single 3m

threshold to eliminate erroneous predictions of shrubs and bare ground as trees. This occurred in open sites, especially at the edges of marshy habitat, where the test of bare ground can appear similar to tree crowns due to low lying vegetation with sharp boundaries against water features. The figure demonstrates the LiDAR threshold for minimum predicted tree height using NEON plot ID OSBS_023 from Ordway Swisher Biological Station, Florida. Using the DeepForest algorithm (B. G. Weinstein et al., 2020a), in blue, several boxes were removed based on no LiDAR returns above 3m (dotted orange). This step was key in open-ground areas in which the algorithm can confuse short vegetation with standing trees.

Figure 4 – figure supplement 2. Illustration and discussion of attempts to incorporate LiDAR into DeepForest algorithm.



During the training of the DeepForest model, we experimented with many approaches to include multiple sources of RGB and LiDAR information on tree presence and crown shape. LiDAR data provides information on canopy height and is therefore commonly used as the primary data source for crown delineation. We attempted to incorporate LiDAR directly into DeepForest by using NEON's canopy height model (CHM) as raster image alongside the three band RGB image to create a four-band stack. While this seems straightforward it changes the model

weights so that they are no longer related to the 3 band pretraining using the ImageNet dataset. This means that the neural network must be instantiated with random weights. This requires the network to learn basic feature representations of colors, shapes, and lines rather than inheriting this kind of information from the pretrained model. Given the paucity of labeled data, this leads to overfitting of the network. For example, when tested on the 18 hand annotated plots for the TEAK NEON site, the RGB-only mAP score was 0.529, whereas the combined RGB + CHM score was 0.389. Informal testing on flat-topped canopies in forests such as MLBS and LENO showed a disconnect between the LiDAR prediction and RGB predictions. In general, the CHM+RGB stack did not offer meaningful improvements in predictions and is therefore not currently included in the model. This may be due in part to the point densities of much of NEON's LiDAR data. Most LiDAR based methods are evaluated on LiDAR data from a single forest type with point densities ranging from ~15 pts/m (e.g., Duncanson et al. 2018) to over 100 pts/m (e.g., Aubry-Kientz et al. 2019). In contrast, NEON's current continental airborne LiDAR program produces ~6 pts/m and these densities can be highly variable, with large areas containing less than 4 pts/m. This tradeoff is needed to cover the much larger areas than previous studies.

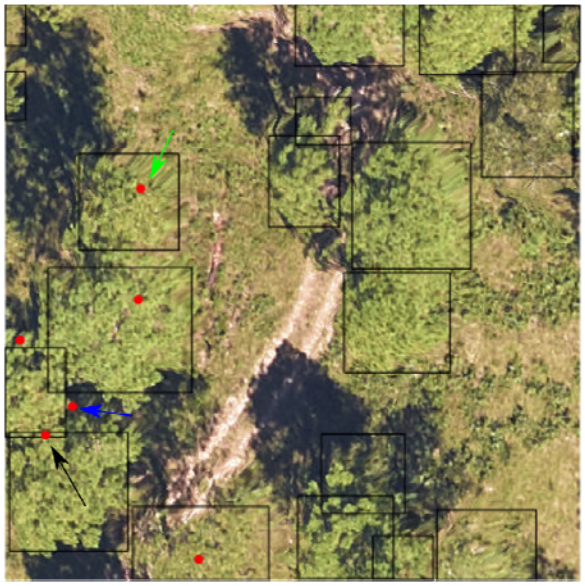


Figure 5 – figure supplement 1. Description of workflow for filtering data to only include live trees likely to be visible in the canopy. Raw NEON tree stem data was processed from the NEON data portal to provide a data set to compare image derived heights to field measured heights. The following filters were applied to the raw NEON field data (ID: DP1.10098.001) after download. A reference tree must have, 1) Valid spatial coordinates, 2) A unique height measurement per sampling period. Individuals double recorded but with different heights were discarded, 3) Height measurements in more than one year to verify height measurement, 4) Changes in between-year field heights of less than 6m, 5) A classification as alive, 6) A minimum height of 3m to match the threshold in the remote sensing workflow. 7) Be at least within 5m of the canopy as measured by the LiDAR height model extracted at the stem location. This was used to prevent matching with understory trees in the event that overstory trees were eliminated due to failing in one of the above conditions, or not sampled by NEON. To match trees in the field and the NEON Crowns dataset, we took the closest height when two predictions and field stems overlapped. We also dropped CLBJ since only 3 points met these

criteria. All other NEON sites did not have any data that met these criteria. In this figure we show stem recall for evaluation plot JERC_048 from Jones Ecological Resource Center, Georgia. Predicted tree bounding boxes in black. Filtered NEON field stems (see above for filtering criteria) in red. The green arrow shows the simplest scenario, a single prediction unambiguously overlaps a single collected stem. The black arrow shows a moderately challenging scenario in which a visually unambiguous crown only barely matches the field collected stem. This could occur due 1) the stem growing at an angle leading to a spatial mismatch between crown and stem, 2) spatial error in measuring the crown location, 3) spatial error in the georeferencing of the RGB image. The blue arrow shows a stem point that does not overlap with a prediction crown. We have written the stem recall evaluation to be conservative, allowing no tolerance for points outside of the prediction box. Therefore, the stem recall for this image was 4/6 = 66.66%, which we believe is a conservative representation of the performance of the algorithm given the uncertainty in the field data and matching process.

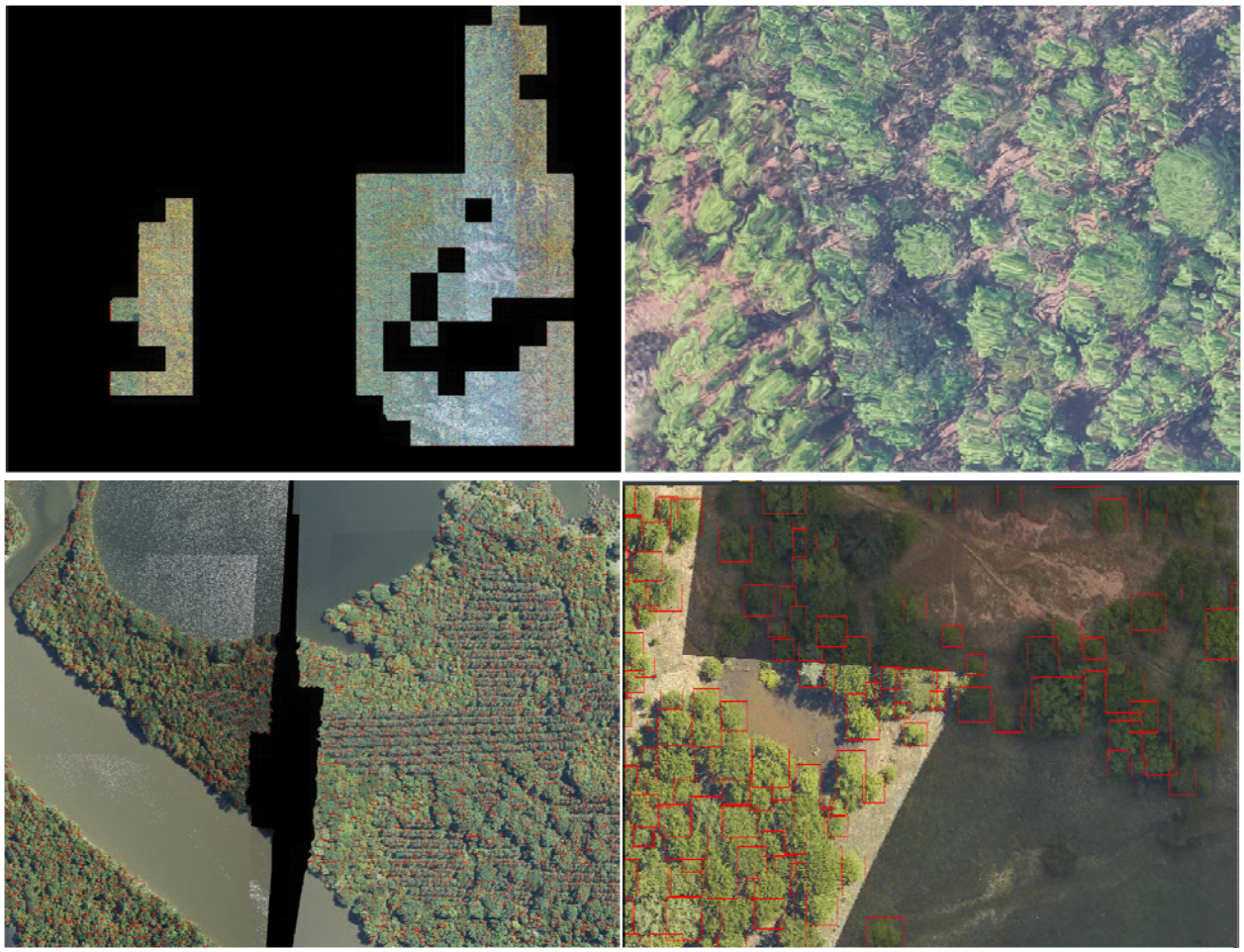


Figure 7 - figure supplement 1. Illustrations of data quality challenges in remote sensing data. With over 7,000 1km2 tiles, it is not possible to do a systematic check of data and predicted crowns. However, there are several general issues that we have encountered that are worth noting. Data quality errors include incomplete site coverage, image artifacts due to georectification, errors in orthomosaic stitching and lighting changes among data collection events. Top left) Patchy coverage at the site level can prevent broad scale analysis with large gaps in RGB data availability (NEON site GRSM). Top right) RGB artifacts at the edge of images stem from the challenge of georectification of the RGB and LiDAR tiles leading to swirling in the RGB pixels (NEON site OSBS). Bottom left) Seams in the flight-lines in the RGB

images leads to gaps in local predictions (NEON site DELA). Bottom right) Changes in illumination among data acquisition collections lead to stark changes in pixel values.

Appendix 1

NEON Site Abbreviations

Site Name	siteID	Domain Number	State	Latitude	Longitude
Abby Road	ABBY	D16	WA	45.76243	-122.33033
Bartlett Experimental Forest	BART	D01	NH	44.06388	-71.28731
Blandy Experimental Farm	BLAN	D02	VA	39.06026	-78.07164
Caribou-Poker Creeks Research Watershed	BONA	D19	AK	65.15401	-147.50258
LBJ National Grassland	CLBJ	D11	TX	33.40123	-97.57
Rio Cupeyes	CUPE	D04	PR	18.11352	-66.98676
Delta Junction	DEJU	D19	AK	63.88112	-145.75136
Dead Lake	DELA	D08	AL	32.54172	-87.80389
Disney Wilderness	DSNY	D03	FL	28.12504	-81.4362

Preserve					
Guanica Forest	GUAN	D04	PR	17.96955	-66.8687
Harvard Forest	HARV	D01	MA	42.5369	-72.17266
Healy	HEAL	D19	AK	63.87569	-149.21334
Lower Hop Brook	HOPB	D01	MA	42.47179	-72.32963
Jones Ecological Research Center	JERC	D03	GA	31.19484	-84.46861
Jornada LTER	JORN	D14	NM	32.59068	-106.84254
Konza Prairie Biological Station	KONZ	D06	KS	39.10077	-96.56309
Lajas Experimental Station	LAJA	D04	PR	18.02125	-67.0769
Lenoir Landing	LENO	D08	AL	31.85388	-88.16122
Mountain Lake Biological Station	MLBS	D07	VA	37.37828	-80.52484
Moab	MOAB	D13	UT	38.24833	-109.38827

Niwot Ridge Mountain Research Station	NIWO	D13	CO	40.05425	-105.58237
Northern Great Plains Research Laboratory	NOGP	D09	ND	46.76972	-100.91535
Klemme Range Research Station	OAES	D11	OK	35.41059	-99.05879
Ordway-Swisher Biological Station	OSBS	D03	FL	29.68927	-81.99343
Red Butte Creek	REDB	D15	UT	40.78374	-111.79765
Rocky Mountain National Park, CASTNET	RMNP	D10	CO	40.27591	-105.54592
Smithsonian Conservation Biology Institute	SCBI	D02	VA	38.89292	-78.1395
Smithsonian Environmental Research Center	SERC	D02	MD	38.89008	-76.56001
San Joaquin Experimental Range	SJER	D17	CA	37.10878	-119.73228

Soaproot Saddle	SOAP	D17	CA	37.03337	-119.26219
Santa Rita Experimental Range	SRER	D14	AZ	31.91068	-110.83549
Talladega National Forest	TALL	D08	AL	32.95046	-87.39327
Lower Teakettle	TEAK	D17	CA	37.00583	-119.00602
West St Louis Creek	WLOU	D13	CO	39.89137	-105.9154
Woodworth	WOOD	D09	ND	47.12823	-99.24136
Wind River Experimental Forest	WREF	D16	WA	45.82049	-121.95191
Yellowstone Northern Range (Frog Rock)	YELL	D12	WY	44.95348	-110.53914